DOCUMENT RESUME

ED 245 297 EA 016 802

AUTHOR	Barber, Larry
TITLE	Teacher Evaluation and Merit Pay. A Background Paper.
INSTITUTION	Education Commission of the States, Denver, CO. Task Force on Education for Economic Growth.
REPORT NO	ECS-WP-TF-83-5
PUB DATE	83
NOTE	39p.
AVAILABLE FROM	ECS Distribution Center, Education Commission of the States, 1860 Lincoln Street, Suite 300, Denver, CO 80295 ($3.00).
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; *Faculty Evaluation; Formative Evaluation; Merit Pay; Professional Development; Salaries; Summative Evaluation; Teacher Effectiveness; *Teacher Evaluation; *Teacher Promotion

ABSTRACT

To make merit-pay plans work, a district's teacher-evaluation system must be both summative and formative--evaluating both for teacher outcomes and for progressive improvement. Only peer-mediated self-appraisal systems seem able to meet such a challenge. Fairly rewarded merit-pay plans depend on accurate teacher evaluations. However, teacher evaluation systems suffer from a confusion of purpose, split between those that only reward or punish teachers (summative) and those designed to improve teachers' performances as they teach (formative). There is a need for a unified system. Research shows that attempting to improve teacher performance through a reward-punishment system alone is nearly always unsuccessful. Furthermore, after falsely labeled merit-pay plans are disregarded, there are clearly few real merit plans, and those most often fail because of inadequate evaluation systems. Often, other problems are also evident in such plans; notably, teacher perception of evaluation plans as aversive control, turnover of administrative staff and school board members, and subjectivity in indexing the qualities of good teaching. Peer-mediated self-appraisal is both summative and formative, avoiding the other pitfalls of evaluation: it is summative for beginning teachers, providing a standard competency level based on peer review; it includes formative features at the probationary level; and it provides a formative professional review on a regular basis for senior teachers. (JW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TEACHER EVALUATION AND MERIT PAY

by

Larry Barber

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. V. Allen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EA 016 802

TEACHER EVALUATION AND MERIT PAY

by

Larry Barber

Working Paper No. TF-83-5

Background Papers for the
Task Force on Education for Economic Growth

Education Commission of the States
1860 Lincoln Street
Denver, Colorado 80295

Additional copies of this paper can be purchased from
the ECS Distribution Center.

CONTENTS

INTRODUCTION 1

1. TEACHER EVALUATION AND MERIT PAY 3

2. THE NATURE OF MERIT PAY 5

3. THE NATURE OF EVALUATION 7

4. THE NATURE OF TEACHER EVALUATION 11

5. REVIEW OF THE LITERATURE: A SEARCH FOR SUCCESS 13

6. MERIT PAY IN THEORY AND PRACTICE 17

7. TEACHERS' WANTS AND ADMINISTRATORS' NEEDS 19

8. RECONCILING THEORY AND PRACTICE IN EVALUATION 21

9. ADDITIONAL PROBLEMS WITH EVALUATION AND MERIT PAY PLANS 23
Aversive Control 23
Difficulty Level of the Task 24
Board and Administration Turnover 24
Nepotism 26
Professional Nepotism 26
Low Teacher Morale 27
Sabotage 27
Absence of Objectivity 28

10. THE NATURE OF PEER-MEDIATED SELF APPRAISAL 35
New and Beginning Teachers 35
Standard Formative Evaluation System 35
Standard Professional Review 37
Remedial Contract and Evaluation Curtain 39
Intensive Assistance 39
Assumptions 40

REFERENCES 41

INTRODUCTION

Teacher evaluation systems seem to be reducible to one of two types -- either systems that reward and/or punish or systems that are designed to improve teaching performance. It appears that successful teacher evaluation systems do not mix these two components: they do either one or the other. There also appears to be an inherent paradox in the whole concept of teacher evaluation. Gene Glass, in an article in The Educational Researcher, suggests the paradox paraphrased here. Excellent teaching appears to emerge when teachers feel that they are not being judged and evaluated. And yet, it appears that the teachers who approach excellence in their profession are the ones who seek out and respond to an evaluation of their teaching (Glass, 1975).

1. TEACHER EVALUATION AND MERIT PAY

Recently three major works have been published that make various recommendations about several aspects of American education. The reports -- A Nation at Risk by the National Commission on Excellence in Education, Making the Grade by the Twentieth Century Fund Task Force, and Action for Excellence by the Education Commission of the States' Task Force on Education for Economic Growth -- focus upon the identification of teaching excellence and in some way suggest providing differential pay incentives for successful teaching performance. Each report, in its own way, suggests the improvement of teacher evaluation systems and suggests forms of differential awards, i.e., "merit pay" for excellence in teaching.

Merit pay, by definition, is a system of compensation for work where payment is dependent upon the output of the individual. The compensation may be in addition to a regular salary or independent of it; or payment may occur in the form of raises granted to the individual earlier than other workers in the same category. In any case, the compensation system is developed around some objective evaluation system wherein the payee's performance is judged to be superior to that of fellow workers.

Merit pay is and has been a popular notion in contemporary American society. The provision of a greater income than average for those who turn in a superior performance is clearly tied to American conservative idealism. Further, merit pay for teachers is tied to the current mythology that business, industry and government, and therefore just about every agency in America, has some form of merit pay system where the best performers obtain greater financial rewards. Mythology, of course, is not reality. Very few industry positions are accompanied by a form of merit pay except those held by people whose financial condition enables them to take advantage of various financial manipulations of the industry. The government certainly does not have merit pay systems in force. One government attempt at merit pay failed miserably (Silverman, 1983). It seems that if merit pay were such a good idea for teachers, federal and state governments would lead the way and show education how to accomplish successful merit pay systems for its own employees.

The current cry for merit pay for teachers is not new. There was a similar concern in 1953-54 and again in 1966-68. This current demand, however, has different concerns than the earlier ones. It is tied to a desire for career ladders for teachers and a need to reduce the number of practicing school teachers as enrollments decline. A prevailing attitude is, "Since we have to let a few teachers go, let's figure out a way to keep only the best."

2. THE NATURE OF MERIT PAY

The nature of merit pay is relatively simple. It is a process or a system that allows for additional compensation for those in a social system who turn in the best performances, produce the most or in some way are the most effective at the common task. Generally only a select few in the system are rewarded, hence the remainder (usually the majority) are, in their opinion, punished. Merit pay systems are generally those that set up a competition for compensation among individuals in the group. Thus it is very difficult to use merit pay systems where group cooperation is desired (such as an elementary school) to accomplish a given end.

A major component of any differential pay or merit pay plan is an objective evaluation system. If you are teaching someone to shoot baskets or pass a football, the evaluation system is relatively easy. You count the frequency of correct hits and reward or punish accordingly. As human behavior becomes more complicated, the evaluation system must also become more complicated. The act of teaching becomes extremely complex and difficult to evaluate and, as we shall see, the systems used to evaluate teachers have not been developed with a sufficient degree of completeness to warrant use for definitive decisions such as retention, promotion, dismissal or merit pay.

Merit pay, as a concept, has not fared well in education, business, industry or for that matter in any social system. At first glance it appears that merit pay, as a concept, is destined to fail. There are a few successful merit pay systems, usually in wealthy school systems, with very teacher evaluation strategies that allow objective, data-based decisions about the teachers under review. The data suggest that before merit pay systems are initiated, carefully planned teacher evaluation systems must first be installed. The culprit in failure may be the evaluation system, not the concept of merit pay.

3. THE NATURE OF EVALUATION

The nature of evaluation is basically the development of a process that (1) judges worth, (2) measures the degree to which specified tasks are accomplished or (3) provides assistance in the developmental process.

In program evaluation we find two basic divisions, formative and summative, as functions of evaluation and categories of evaluation research (Anderson et al., 1975). Formative evaluation essentially is concerned with helping the developer of programs or products such as curricula, books, etc., through the use of empirical research methodology in the process of the development and implementation of the program. Summative evaluation is concerned with evaluating the overall program after it is in operation or after it has been completed. Ideally the formative evaluator will work along with the developer from the outset of his work at the same time the development planning is taking place. Summative evaluation strategies deal with after-the-effect evaluation. Please note that we are not attempting to assume that formative evaluations are exempt from scientific rigor, rather only that they may be simplified, more error-prone versions of later summative evaluations. An important feature of the formative evaluation system as distinct from summative evaluation is that formative must occur in close collaboration with program or product development. The formative evaluator is a part of the development process and it is not unreasonable for that person to become as ego-involved in the program as his scientific demeanor will allow. A summative evaluator, on the other hand, should be and should seem to be independent of the developer product or program. Thus, the person who carries out the formative evaluation may be of considerable help to the summative evaluator in terms of the problems of measurement, etc., but the formative evaluator should not ever become the summative evaluator. "Summative evaluation, given adequate time and funds, can and should be technically excellent research, in that summative means assessing the impact of a program" (Anderson et al., 1975).

If these concepts of summative and formative evaluation are applied to personnel evaluation, we can more clearly see the distinctions between the two components. Reward and/or punishment evaluation systems are clearly summative evaluation in the sense that the identification of effective teaching behaviors or the identification of teacher competency and teacher behavior can be looked at and judged through the eyes of another person, the evaluator. On the other hand, teacher evaluation systems designed to allow individuals to improve their teaching performance are clearly formative in nature.

One of the things that we know clearly is that if we assign a formative evaluator to a developing program, the evaluator is likely to become a part of the developing project. In essence,

the formative evaluator becomes the project auditor who, with the goals and objectives of the funding agency or the initial purpose of the developing program, keeps trying to bring the developers "back on track," while providing them with as much objective data as possible to accomplish their initial task. We also know that formative evaluators begin to lose objectivity and begin to identify clearly with the developing project even though their role is both "objective and evaluative." Thus, we almost never assign a single person the role of being a formative evaluator of a project and then later the summative evaluator. Perhaps the same thing should be true of the different forms of teacher evaluation, and this may be why it is nearly impossible to mix successfully a teacher evaluation system that both provides reward and/or punishment and is designed to improve teaching performance.

Perhaps we need to develop two forms of evaluation. One form of evaluation would be formative and designed to improve teaching performance. It could be applied to most of the teachers in a school district. At least 95% of the teachers currently employed are excellent, they do their job well, they understand it, but they feel they need assistance in improving their own teaching performance on a daily, weekly or annual basis. If we were to develop a formative teacher evaluation system, we could assign a formative evaluation person and/or strategy to that process and continually provide feedback to teachers about their effectiveness. For those teachers for whom this process does not work, we would need to switch to a second system that can be either rewarding or punishing -- in essence a summative evaluation process. In so doing we need then to switch evaluators as well. Thus we would place, as an example, all teachers on a formative evaluation system until we find that this system is not working for a very few. These few, according to both the teachers' and the formative evaluator's judgments, would be unable to continue to be productive teachers or learn to be productive teachers on this system. The teacher-supervisor, the teacher and the evaluator then would need to concur that this teacher needs to be switched to another system -- a summative system with a different evaluator. This concept already exists in many teacher evaluation systems and in many districts it is called placing a teacher on intensive assistance, professional aid, review, report, or one of several other terms meaning basically the same thing -- that a teacher now is determined to be less than competent but can be saved if provided with some intensive assistance from another professional.

At this point, to preserve the legal rights of the teacher, we need to introduce the idea of an evaluative curtain. In essence, because of the legal issues of due process, equal consideration under the law, notification of an intent to change the system, etc., we need to draw the curtain on all past data. All things that have preceded the curtain cannot now be used for or against that teacher in making judgments after the curtain has been drawn. All records need to be expunged, all data needs to be stopped and

the teacher now, in essence, gets a fresh new start with a new evaluator, a new form of instruction, intensive assistance or inservice training, and a data collection system designed to provide feedback that will make the teacher competent. But all parties should be made clearly aware that if the teacher, with assistance from others, does not reach a competence level that satisfies the employer then this new data system, developed after the curtain has been drawn, may be used to terminate the teacher. We will explore these concepts later.

It also seems important here to point out an inherent difference between program evaluation systems and personnel evaluation systems. Program evaluation since the early seventies has approached something of a science in that, through program evaluation methodologies and procedures, one is able to very carefully determine whether a program is being implemented and whether it is accomplishing those things that it says it will do. The same level of development has not appeared in personnel evaluation systems. Scriven makes this point clearly.

> Teacher evaluation is a disaster. The practices are shoddy and the principles are unclear. Recent work has suggested some ways to clarify the issues and to make the procedures more equitable and reasonably valid, but one cannot yet point to a single exemplary system in which the practices come near to matching our knowledge. This state of affairs is a terrible indictment of the universities, since the problems are sufficiently tractable that a modest investment of only 2%-3% of their internally controlled research funding would have solved them 50 years ago. That the great research and teaching centers of learning never put serious efforts into research in teaching, though they had to evaluate teachers all the time, is one more scandal in the history of the guilds of professions that fail to practice what they profess -- in this case, research-based decision making. Once more, the pressure to reform is coming from outside, from the Congress and the courts, not the conscience (Scriven, 1981).

4. THE NATURE OF TEACHER EVALUATION

It is clear that we all struggle with the concept of teacher evaluation, but what does it mean? What is this concept that has taken so much of education's time and energy over the past few decades? Max Abbott, in a 1975 monograph, suggests the following definition. "The results of personnel assessment should be made available to the individual as a basis for personal and professional growth and improvement." A further definition suggests that, "Teacher evaluation exists for (1) improving of teaching through the identification of ways to change teaching systems, teaching environments or teaching behaviors; (2) supplying information that will lead to the modification of teacher assignment, such as placements into other positions, promotions and terminations; (3) protecting students from incompetence and teachers themselves from unprofessional administrators; (4) rewarding superior performance; (5) validating the school system's teacher selection process; and (6) providing a basis for teachers' career planning and professional development" (Abbott, 1975). Others have suggested that the purpose of a personnel system is to provide assistance to teachers to improve teaching performance in the classroom and in preparation for the classroom. However, that is not the way all people see it. Recently a school board member defined his intent for the development of a teacher evaluation system as follows. "I want something that we can use to get rid of incompetent teachers."

The problem is that too many people ask too many things of educational personnel evaluation systems, and too few understand the philosophical ramifications of those differences. For example, G.B. Redfern suggests that, "To believe that producing a teacher evaluation system for the purpose of guaranteeing that children will learn what is taught is nonsense" (Redfern, 1980). What is important to understand at this point is that teacher evaluation can be reduced to two components, either systems that reward and/or punish (merit pay systems), or systems that are designed to improve teaching performance. The two components have not been mixed effectively, for the most part. In the review of the literature that will follow, we will try to point out that the research reports clearly show that any time anyone attempts to improve teaching performance through a reward and/or punishment philosophy, the system, with very few exceptions, is almost inevitably bound to fail.

Much of the literature about personnel evaluation has been at best misleading and at worse erroneous, suggesting that a single evaluation system that is at once summative and formative can be implemented by a single person, usually the school principal. To explore this confusion, let's retreat to the Abbott definition of evaluation and quote the whole paragraph. "I want to emphasize at the outset that there are two major reasons for assessing (evaluating if you will) performance. First and most important,

the results of assessment should be made available to the individual involved, to be used as a basis for personal and professional growth and improvement" (note: a clear formative evaluation strategy). Abbott continues, "Inevitably, however, decisions must be made regarding such issues as promotion, retention, and salary improvement. Thus, assessment data also serve a managerial function, that of providing a rational basis for making those decisions" (note: a clear summative evaluation strategy) (Abbott, 1975).

Abbott was talking about administrative evaluation systems, but the point is the same. He and most other authors suggest that we use one system that is both formative and summative and is managed by one person.

Lest the reader think that we are picking on Abbott, we hasten to point out that confusion exists in almost every piece of literature about personnel evaluation. The message is clear -- develop a single system that is both formative and summative and let it be managed by a single person, the teacher's supervisor.

5. REVIEW OF THE LITERATURE: A SEARCH FOR SUCCESS

In preparation for this paper, the author has reviewed literally hundreds of papers, research articles and books on teacher evaluation, and found very few examples of successful practice in either the development of teacher evaluation systems or merit pay systems. I have developed an arbitrary definition of "successful" evaluation systems. Criteria for "successful" were:

1. The evaluation or merit pay system in practice actually does what it says in its written description

2. The system, as described in writing and operationalized in practice, has been installed and operated for at least two school years.

3. Spokesmen for administration and board and the teachers' association or union agree that the system is functioning as written and is beneficial to children.

Educators constantly tinker with personnel systems, adjusting a part here or there, and inevitably the operating system is very different from its written description. In judging whether or not personnel systems in fact accomplish what they set out to accomplish, one must examine them in the light of program evaluation techniques.

When I first began searching for successful teacher evaluation systems or successful merit pay plans, I was informed of scores of successful programs. "Successful," that is, until I read the description or talked with the implementors. People said they had a successful merit pay plan and then described a plan in which compensation is dependent solely on input variables (number of credit hours, type of courses taken, etc.) or on longevity (i.e., the longer a teacher teaches the more merit they have). Of the scores of merit pay plans we have followed up, only a few (perhaps 10) are really merit pay plans. Most are something else.

There is a great deal of confusion in the areas of teacher evaluation/merit pay. After virtually hundreds of phone calls and reading hundreds of pieces of literature about teacher evaluation in an attempt to understand the issues surrounding it, one comes away with several somewhat schizophrenic thoughts:

1. There does not appear to be any consistent systematic development of teacher evaluation systems by any agency that should be responsible for the development of those systems. Phi Delta Kappa's CEDR center has tried over the past seven years to coordinate the development of teacher evaluation systems, strategies and methodologies, but thus far has not

produced exemplary products suggesting how such systems should be developed.

2. It seems that, even though it is possible to develop a successful teacher evaluation system, every school system in the country will have to do it in its own way. No one seems to be able to take a successful system developed in one place and apply it intact to another school district. In essence, each district will have to negotiate the terms in either contract form or some other format.

3. It seems that only a very few (10 to 15) teacher evaluation systems have been developed that have a reward and/or punishment component and/or a merit pay component that has lasted for two years without major alterations.

4. Teacher evaluation systems seem to be promoted by educators to keep the public or the state happy. If left alone, educators develop teacher evaluation systems that assist teachers in self improvement, not merit pay plans.

5. Teacher evaluation to assist teachers in learning new skills is obviously a good idea. Attempting to include punitive or even rewarding components into a system designed to improve teaching performance seems inevitably (with few exceptions) to cause the system to fail.

6. It is far more expensive and time consuming to develop a personnel evaluation system than almost anyone realized when they started out to do so.

These findings are similar to the results of an earlier study conducted by Glen Robinson. In that extremely comprehensive study, Robinson examined the effects of merit pay plans upon 239 districts from 1908 to 1978 (Robinson, 1979). He found the following kinds of reasons for the failure of merit pay plans (Robinson, 1983):

> Forty percent of the districts stated that the administration of merit pay forced the termination of the plan. By that they mean (1) plans were too complicated to be administered fairly; (2) plans were too much of a burden on a limited number of administrators; (3) plans made no difference in teaching performance; (4) there was too much record keeping required; (5) as administration or leadership changed, so did plans; (6) changes in school boards brought changes in plans; and (7) parents wanted children taught only by superior teachers.
>
> Thirty-eight percent of the districts where merit pay failed stated that a merit pay plan created personnel problems of sufficient magnitude to terminate the plan:

(1) teacher morale suffered; (2) the plan created jealousy and friction among staff members; (3) staff charged administration with unjustified favoritism for a few teachers; (4) the plans destroyed cooperative team work and produced competition; and (5) teachers' organizations and unions opposed merit pay.

Seventeen percent of the districts with failing merit pay plans gave these financial problems as the primary reason they abandoned merit pay plans: (1) lack of funds to administer the plan; (2) incentives too low to bring about changes in teaching performance; (3) not enough teachers rewarded to make the plan work; and (4) collective bargaining negotiated the plan out of the budget.

Other problems were: (1) evaluation procedures were unsatisfactory; (2) it was very difficult to determine objectively who deserved merit pay; (3) there was not enough data developed by the plan to defend the choices made, hence charges of arbitrariness were filed; (4) evaluators were inconsistent in their ratings; (5) evaluators were not able to be objective, thus they were subjectively partial to a few people; (6) no satisfactory evaluation instruments were available; (7) teachers charged evaluators with unjustified subjectivity; (8) arbitrary cutoff points seemed illogical and the quota system froze out younger members; and (9) it was impossible to find superior results, or in fact to measure any results.

Time and time again the implications from Robinson's study, and from our search after his study; suggest that the major fault of merit pay plans is inadequate evaluation systems -- those that were subjective or poorly thought out or administered.

6. MERIT PAY IN THEORY AND PRACTICE

Popular discussion of merit pay programs has termed a variety of programs "merit pay," many of which are something other than merit pay. For the purposes of this paper, the term "merit pay" will be used for those programs that make awards over and above a regularly received salary; or that place teachers on accelerated pay schedules, and that grant awards as the result of evaluations that judge the efforts of teachers to be superior. Those programs that award additions to base salaries without judging the quality of efforts are considered "incentive programs." Therefore, teachers who receive cash awards for accepting assignments in inner-city schools are not participating in merit pay programs, but in incentive plans.

Merit pay's track record has been very problematic in education (Robinson, 1983), in industry (Brinks, 1980) and in the federal employment system (Silverman, 1983). Many of the problems cited by sources outside of education are the same ones that plague merit programs in schools. But from the standpoint of accepted personnel practice, merit pay is particularly inappropriate for educators. James T. Brinks (1980), an expert on compensation, benefit and performance measurement systems, points to three conditions that must exist before a merit pay plan can motivate workers to be more productive. The first precondition is that existing salaries should be externally competitive. Second, salaries should be internally equitable by reflecting the employee's appropriate level of responsibility, years of experience and the teacher's sustained performance level. Unfortunately, in the majority of cases, salaries in education (especially for teachers) are neither externally competitive (with business and industry) nor internally equitable (given the extended plateau in pay schedules for experienced teachers). A third critical point, one that will be addressed again later, is that there must be objective work measures. Brinks (1980) points out that it is difficult and expensive to create objective measurement systems.

Collecting case studies of failed merit pay systems is not difficult. Educational Research Service's twin studies (1979) on merit pay for school administrators and merit pay for teachers document thoroughly the small percentage of programs attempted and the even smaller percentage of those that survived. Examples continue to appear from school systems around the country: Brewster, New York (Monk, 1983) and Kalamazoo, Michigan (Doremus, 1982) have been reported on within the past year. A decade ago, Robert Bhaerman, while serving as director of education research for the American Federation of Teachers, documented a history of merit pay systems (Bhaerman, 1973) which includes a four-year feasibility study conducted in North Carolina in the early sixties that rejects a statewide merit pay plan as neither practical nor feasible.

Historical views are interesting and instructive, but recent developments in education's management should not be ignored in considering the current popular support for merit pay. Declining enrollments have created a host of pressures to produce the same or more benefits with dwindling resources. Merit pay has been reintroduced as a means of spurring on teachers and administrators. Consider the impact of an evaluation system that delivers merit pay for a few, no merit pay for most and the possibility of termination for those not judged meritorious. Even when evaluation systems for reduction-in-force considerations and those put in place to decide merit are separate, teachers are sensitive to the potential for abuse of information.

7. TEACHERS' WANTS AND ADMINISTRATORS' NEEDS

Whether to evaluate is not the question. Administrators are called upon to make decisions concerning retention, promotion and termination. Making these decisions based on defensible data is the only sane path in the face of potential lawsuits and demands for accountability. Merit pay introduces another irrefutable demand for evaluation. Teachers want to know how well they are performing -- for their own growth as well as job protection.

Traditionally, the same evaluation system has been called upon to deliver data to satisfy both parties, despite useful distinctions between the formative and summative processes (Scriven, 1973).

8. RECONCILING THEORY AND PRACTICE IN EVALUATION

Scriven considers the evaluation process to be the same, whether teachers or curricula are the object of evaluation. In terms of goals, evaluation attempts to answer certain types of questions about certain educational instruments like processes, personnel, procedures and programs (Scriven, 1973).

Treating teachers as an "entity" becomes unworkable when we try to apply formative and summative methodology. The wants of teachers include the outcomes of formative evaluation. Scriven describes these outcomes as the evaluation feedback loop. In this the formative evaluation stays within the development agency and serves to improve the product. The methodology used includes self and peer evaluation and small group testing with reworking the product. Summative evaluation, by outside experts, decides the worth of the finished product. The results of this form of evaluation are intended for public dissemination. What is the "finished" product of teaching?

The answer that many administrators have reached, for their own needs and to implement merit pay plans, has been to adopt end-of-the-year gains of nationally normed achievement tests as "outcome" measures. Teachers and professional evaluators raised vigorous objections to this practice because several uncontrollable factors (knowledge level of incoming students, home environment and the different ability levels of students) influence the scores. Regression error and the limitations of achievement tests are sources of misinformation that penalize teachers when achievement score gains are treated as "final products" of teaching (Soar et al., 1983).

9. ADDITIONAL PROBLEMS WITH EVALUATION AND MERIT PAY PLANS

As we examined teacher evaluation and merit pay plans, we noted problems almost identical to those found in Robinson's 1979 study. In addition we contacted a number of the people who were in charge of the administration of these plans and asked them to tell us the reason why they were having problems. Some of these reasons were different than those Robinson had reported. A summary of the reasons for the failure of merit pay and evaluation plans, that have a high reward or a high punitive function, follows.

Aversive Control

The most significant reason for the failure of these plans seems to be a concept that we will call the perceived use of aversive control. Designers of evaluation plans insist they do not have aversive control in their evaluation systems on the grounds that it won't work. Most teachers and administrators feel that any form of control over their professional lives is aversive. Teacher perceptions of aversive control found in teacher evaluation plans is perhaps the most basic problem of teacher evaluation systems. It is still an unresolved and undiscussed issue. The issue concerns an attempt to control the teacher's behavior in the classroom or the principal in the office. Teachers and principals clearly perceive this attempt at behavior modification or control as extremely aversive and react as anyone does to aversive control -- they create a ministrike.

The following reference, taken from Psychology Today, describes human reaction to aversive control. "During the third Skylab mission in 1974 the flight crew was notably irritable and hostile toward the ground control officials in Houston, who generally responded with verbal reprimands when the crew's performance fell short of expectation. This negative reinforcement eventually resulted in a meeting by the crew about halfway through the 84-day mission. The crew simply stopped working and did what they pleased until mission control agreed to a revised and lighter work schedule." (Author's note: This is the basic concept of a strike.) Having established that aversive control causes increased hostility and diminished productivity, the researcher Amerian stopped using the technique, stating "It was simply too powerful, too destructive" (Rice, 1983). Yet today most teacher evaluation systems are based on aversive control, even though it is clear that aversive control almost invariably produces some form of strike. Little wonder that most teacher evaluation systems are nonproductive and problematic in their administration!

Difficulty Level of the Task

A second major problem, similar but different from that found by Robinson, is that of the whole concept of the difficulty level of the task. The development of a teacher evaluation system is generally something that school administrators have to be forced to do. We have yet to find an administrator who voluntarily began to implement a teacher evaluation plan that is aversive, rewarding or punitive. The force that causes them to produce these kinds of evaluation plans comes in one of several ways. Either the state legislature passes a law, the school board chooses to get rid of a little deadwood or the public demands that somehow public employees be held accountable for the expenditure of public funds. Whatever the reason, the force is usually connected with a demand that the superintendent do something he or she does not want to do: develop a rewarding and/or punitive teacher evaluation system.

Consequently, the first attempt by a superintendent at the development of a teacher evaluation system is generally made under duress. Superintendents react by seeing that the new system is underfunded, understaffed and focused on things that are most easily evaluated. The system initially ignores those things that are difficult to evaluate -- motivation, professional pride, creativity, concern for children and many other valuable attributes. These are forsaken for things such as the number of credit hours, longevity, student test scores, number of student contact hours, and all of the rest of those easily counted things that seldom matter in good teaching. Teachers find evaluation leaving a bad taste in their mouths and are unwilling to participate in second or third attempts. By the time the system or superintendent realizes that extensive resources and time must be committed to developing a successful evaluation system it is too late. All the teachers and midlevel administrators in the system have banded together to fight further development.

Board and Administration Turnover

Another problem that affects the development of teacher evaluation systems is the turnover of top administrative staff and board members. Two examples of the effect of turnover are explained below.

In the first example, Superintendent Monk (1983) started with a management-by-objective concept to develop a merit pay plan for the administrators of his 2,500-pupil system. Each administrator had regular responsibilities and also had to select four to six individual objectives from a list the school board provided in addition to regular duties. These objectives were designed to "reflect the needs of the community as the school board interpreted them." As the management-by-objectives program got

under way, several programs improved, all programs seemed better generally, and the school district was perceived as better in the eyes of the community.

"But one problem ultimately undermined the program. The board that had endorsed merit pay gradually turned over and the new board members never came to understand the pains their predecessors had gone through to establish the system. Nor did they fully appreciate the reasons for the many positive changes the system had seen in the previous years." Thus the program failed. Monk concludes by saying,

> Consequently, merit pay will not work, has not worked, cannot work in the public schools of New York State, nor for that matter in any school system that elects its governing board from among lay persons who are unfamiliar with the internal workings and history of the institution of the public schools.
>
> I believe this reflects an inherent weakness in the public school system in the United States. We cannot expect that an organization of any kind can survive with such a changing, unfamiliar (a harsher but more accurate word might be "uninformed"), publicly scrutinized, frequently second-guessed and voluntary governing body.
>
> The proof of my point: the lack of systems of merit pay in public schools, the lack of stability among school superintendents and middle managers, and the serious erosion of confidence in the public schools throughout the U.S. -- all the results of inconsistent governance and fractionalized interest of frequently elected bosses. It's a grim picture (Monk, 1983).

The second example of the problems caused by the turnover of people in public schools as teacher evaluation or merit pay plans are executed comes from a school district that tried to develop an administrator evaluation system. This development took over five years, involved hundreds of people, got into a major negotiation with the teachers' organization and promoted the development of an administrator's collective bargaining unit. The cost for development, staff time and expenditures was in excess of $200,000. Then it started all over again when a new superintendent refused to implement the plan, but wanted a plan with his own stamp on it. This new system has been in development now for four years and has cost an additional $250,000. As of 1983 it has not been carried out.

Nepotism

As I talked with people about the problems in carrying out teacher evaluation systems, a problem that seems to occur in small districts, but can also be frequently found in larger districts, was revealed. It is the exclusion of a number of people from the merit pay plan because of nepotism. For example, in one district we observed data indicated that at least 35% of the employees were related by birth or marriage. Top administrative staff had brothers, sisters, sons, daughters and wives employed in the district. There were over 80 employees that, in essence, were exempt from any form of evaluation because a relative supervised a person who would evaluate them. One assistant superintendent commented to me, "It's very hard to evaluate two of my principals, even though I'm convinced they're not very good, because they are the daughter and son-in-law of the superintendent who evaluates me."* It seems that any teacher evaluation or merit pay plan might be successful only if every member of the teaching faculty were included in it. When a group of people are excluded for reasons other than competence, the plan is almost always destined to fail.

Professional Nepotism

There is another problem, not noted in the literature but in personal contact with people who have experienced the problems of administration of a teacher evaluation system: professional nepotism. Basically it is the old saw of the "good ol' boys club" or the "good ol' girls club." Many educators develop lifelong friendships either in graduate school or in initial work experience. Some however, have developed long-standing ties to particular professors in schools of education who are powerful in the placement of administrators in that state or in several states in a region. In either case, relatively large coalitions of individuals become bound together with motivations and obligations that are stronger than those in their present job. They protect one another, assist each other in obtaining jobs and help each

---

*Please note, the author does not intend to question the integrity of educators by implying that they selectively hire relatives. Put simply, the children of educators tend also to become educators, and educators tend to marry educators. Hence, in a small community where there is often only one education employer, many of the people eligible for employment will be related. The point here is that, once employed, relatives of upper level administrators are, in fact, exempt from evaluation, and for any evaluation system to work, it appears that no employee should be allowed to be exempt from the system. In cases where people are not exempt from an evaluation system where nepotism is in evidence, such as merit pay systems, it becomes clear who receives merit awards.

other out of poor evaluations. When one of this group becomes a top administrator, friends are exempted from the current evaluation system. For example, an upper-level administrator will chastise a lower-level administrator for giving one of the upper-level administrator's friends a poor evaluation by suggesting that their evaluation must be wrong because "I hired this person and he or she has to be good, so why would you give them a poor evaluation? It reflects negatively on my judgment, and don't forget who evaluates you."

Low Teacher Morale

Another major problem discussed by people implementing teacher evaluation and merit pay plans seems to be that these systems defined as successful by the board, the public or the administration seem to be very disturbing to teachers and produce low morale, which in turn sabotages the teacher evaluation system. The logic now becomes convoluted. When a seemingly workable teacher evaluation system actually is produced, its byproduct is more grievances, confrontations between teachers and evaluators, unrest in teacher association ranks, and coalitions of teachers in separate buildings. As a result of this, teacher morale begins to drop and principals determine that low teacher morale negatively affects the teaching that occurs in classrooms. Eventually, someone begins to tinker with the system to repair low teacher morale. They tinker with it until they change it drastically, until they perceive that morale begins to rise, grievances go down and there are fewer labor relations problems. It seems that a successful teacher evaluation system, as defined by the public and the school board, is one that cannot be tolerated by professional teachers. Those who must implement the system, tinker with it until it is tolerable for teachers, thus it is no longer tolerable for administrators, the public or the board.

Sabotage

Sabotage may destroy even successful evaluation systems. If, by chance, a teacher evaluation system does not negatively affect teacher morale, then the principals (the primary people who carry out the evaluation) often step in to destroy it because, then, administrator evaluation systems cannot be very far behind, and that certainly cannot be tolerated. If teacher evaluation systems can be made nonfunctional and ineffective, then administrator evaluation systems will not be installed.

What our information suggests is that after teacher evaluation systems have been in effect for three or more years, they become generally ineffective -- they simply do not evaluate or make judgments and do not bother anyone. Teachers still may be fired from the district, but not because of poor evaluations. They are terminated because they committed a felony, were caught drinking

on the job, were insubordinate, caught shoplifting, or caught at some other behavior unbecoming a school teacher. The apparent conclusion is that the best teacher evaluation system always seems to fail first, leaving only the ineffective ones.

Absence of Objectivity

A final problem I have noted in conversations with people attempting to implement teacher evaluation or merit pay plans deals with the lack of objectivity in relationship to the measurement of teacher performance. Many systems attempt to develop a defensible index of teaching skills to make comparative judgments about Teacher X compared to mythical scores of optimal performance, or to identify equally mythical lower cutoff scores. This index also identifies minimally effective teacher behavior. If teachers are not up to the minimum, they will be placed on an assistance program. Any teacher who performs at the optimal level is eligible for some form of merit award, or would be identified as a master teacher or given some other superior designation that would provide additional compensation of some sort. In systems that have attempted to do something like this, administrators inform us they are unable to come up with an index of teaching skills that teachers agree to. Educators in general have been searching for a long time for some form of index that could measure teacher competence. "Since the early 1900s, educational researchers have spent endless hours trying to devise a satisfactory measure of teacher competence. In view of the enormous energy expended in trying to identify this elusive prey, some disgruntled researchers have opined that a defensible index of teaching skills ranks third behind two other hard-to-locate objectives -- the Holy Grail and the Fountain of Youth" (Popham, 1975). Popham may have overstated the case a bit, but the point is that developing an index of teaching skills is a difficult task -- one that requires time, thought, people, energy and money.

One problem that has surfaced time and time again is a possible high correlation between the implementation of summative evaluation systems and consequent low morale among teachers. As we examined the problem it looked as though the culprit was summative evaluation itself. But several summative evaluation systems have appeared to be successful. The next most likely candidate for producing low morale seemed to be evaluation systems that combined formative and summative systems or those that said they were formative, but in reality were summative.

Popham (1975), in his chapter on "Teacher Evaluation -- A Special Case," suggests that ". . . there are as many, if not more, dividends to be accrued from the formative evaluation of teachers." Yet he fails to specify why. He then suggests that teachers be considered ". . . as an instructional treatment that with formative evaluation assistance can be markedly improved" (Popham, 1975). Further, Popham suggests that ". . . teachers'

reception to formal evaluation efforts will be far more positive if a full blown formative evaluation program precedes any major summative evaluation effort." He fails, first, to describe more fully the substantive differences between formative and summative evaluation systems as they apply to teachers. A second error is to imply that teachers be considered an instructional treatment that is manipulated and judged by others, while at the same time being evaluated under a formative system. As we shall see, the two are not compatible.

Scriven (1979) suggests the different purposes of formative versus summative evaluation: "Ideal formative evaluation would consist of evaluation with diagnostic and therapeutic recommendations." But diagnostic and therapeutic recommendations are really components of summative evaluations. What is needed is a different -- or clearer -- definition of formative evaluation if it is to be used in a personnel evaluation system that rewards and punishes.

There are few references to the use of formative versus summative evaluation strategies applied to personnel evaluation. Scriven has a chapter in The Handbook of Teacher Evaluation (Jason Millman, ed., 1981) titled "Summative Teacher Evaluation" which introduces "summative" strategies that should be considered formative, thus adding to the confusion about the differences between formative and summative evaluation strategies when applied to personnel evaluation.

Knapp, in an excellent article that helps clarify the problem, suggests that "the primary issue appears to be this: Can schools achieve both formative and summative purposes through more formalized, systematic means? Can these purposes be accomplished within the same evaluation system? The bulk of the writing reflecting the practitioner's view seems to favor separating the two purposes" (Knapp, 1982). That indeed may be the real problem. Systems that suggest they can provide both a formative evaluation (to improve individual teacher performance) and a summative evaluation (to insure system quality) may not do either very well.

It seems a sound idea to take what we know about program evaluation and apply it to personnel evaluation. In program evaluation the differences between formative (process) and summative (product) evaluation are clear, both conceptually and in application. Summative evaluation, as understood and used in program evaluation, seems to be transferable to personnel evaluation with few problems. The problems occur in the attempt to transfer formative evaluation concepts from program evaluation to personnel evaluation. There is an obvious lack of clarity about the transferability of formative systems from program to personnel evaluation, and it appears that one must go beyond evaluation concepts to understand the relevant issues. One must examine the underlying belief systems about the nature of man to

understand the problem of personnel systems, especially those that purport to do both a formative and a summative evaluation.

The problem occurs as the result of a lack of congruence between an evaluator's philosophical belief about the nature of man, his knowledge of evaluation theory, and his application of that theory to a personnel evaluation system. Each of these concepts must be congruent with the other before individuals will cooperate with an evaluation system that purports to help them become better teachers or administrators. Let's explore this congruence theory a bit further.

If an evaluator's underlying philosophy is that man is inherently good, that professional teachers can evaluate themselves better than "outside others," that professionals strive for individual excellence, that data (knowledge) is the best way for the professional to improve himself, but accepts a theory of evaluation that is summative (i.e., rewarding or punitive) where decisions are made about one individual (the evaluatee) by another individual (the evaluator), then there is a lack of congruence in the evaluator's logic. This incongruence is very clear to the evaluatee, and leads directly to both fear and doubt that the evaluator has the evaluatee's best interest at heart.

Lack of congruence occurs when an evaluator (or an evaluation system) has a philosophical belief system (stated above) that is formative in nature, follows with a theory of evaluation that is also formative, but then attempts to implement a summative system, either alone or in combination with a formative system, as though the two systems were one. Lack of congruence can also occur when the beliefs are summative, the theory is formative and the implementation plan is summative. Congruence occurs when the philosophical-theoretical and implementation systems are consistent. The evaluatee at least understands what is to happen to him and why. Even though he may not like summative systems, he at least understands the motives behind them and will accept summative evaluation systems far more readily when the evaluator's philosophy, theory and implementation plans are congruent. What is important in the eyes of the person being evaluated is that the system of evaluation must be congruent with the stated purposes of evaluation. If I am a teacher and my evaluator says that the purpose of evaluation is to improve individual teacher performance, yet evaluates me under a system that requires me to produce data that can be used against me, then I am going to think he really does not believe his stated purpose of evaluation and further, I am not going to believe anything else he says.

From the hundreds of examples of teacher evaluation systems found in practice, it seems that the systems that have succeeded the best are those where it is clear from both the teachers' and administrators' points of view that the intent of the teacher evaluation system is to provide information to an individual

Figure 1. Congruence of Philosophy. Theory and Practice in Evaluation

Formative Evaluation	Summative Evaluation
Philosophy:	
Each individual strives for excellence.	Individuals achieve excellence only if supervised or evaluated by others.
Theory:	
Evaluations done to improve the performance of the individual. Reward or punishment should be decided internally.	Evaluation is done to improve the performance of the social system. Reward or punishment should be decided externally.
Practice:	
Evaluate the process of instruction but not the person.	Evaluate the products of instruction as well as the process and the person.

teacher in such a manner that only they can own and use that information to improve their own teaching performance. For the vast majority of teachers, this is a fine system. They are competent, caring, kind and interested educators attempting to do a better job daily than they have done the day before. The kind of system where people are allowed to make their own judgments about how they improve themselves is effective for professionals. At least 90% - 95% of the teachers in a district are able to deal effectively with teacher evaluation systems such as this and encourage and assist its implementation, and as a result become better teachers. You can have one system that is designed to improve teacher performance for the vast majority of people and a different system for those who need assistance improving their teaching performance.

Successful formative systems can be found where the greatest academic freedom occurs -- universities. These systems generally encompass self assessment, peer review or various combinations of both.

Gregory J. Carroll defines a system of self evaluation as enhancing self understanding and improving instruction. He suggests five major aids for self evaluation and gives examples of each: (1) self rating forms, (2) self reports, (3) self study materials, (4) observation of colleagues' teaching and (5) video tape and audio tape feedback of one's own teaching (Carroll, 1981). Carroll suggests a model for self evaluation called a self confrontation model that describes and defines microteaching, interpersonal process recall and interaction analysis.

In another study, Grace Lazovic suggests that the essential function of a peer review is to guarantee that members of an academic community have control over their own standards, membership and the future course of their discipline (Lazovic, 1981). Lazovic further suggests that faculty peers are uniquely qualified to judge the substance of teaching because faculty possess two distinct requirements for the judgment task: "(1) their knowledge of the discipline being taught provides a background against which comparison can occur, and (2) their long training in the evaluation of evidence enables them to weigh what is revealed through documentation." Lazovic suggests five questions that should be considered when peers review each other for evaluation purposes.

1. What is the quality of the materials used in teaching?

2. What kinds of intellectual tasks were set by the teacher for the students and how did the students perform?

3. How knowledgeable is this faculty member in the subjects taught?

4. Has this faculty member assumed responsibility related to the department or university's teaching function or mission?

5. To what extent is this faculty member striving for excellence in teaching?

Lazovic also describes and clarifies the peer review process, provides peer review forms and describes a model peer evaluation system. It is important, however, to note here that the concepts that Lazovic introduces in peer review and the criteria that others look at in reviewing their peers are very similar to the suggestions that came out of effective schools research. Those suggestions are almost identical to Lazovic's in that they suggest that teachers know their subject matter, that they care about their children, that they are concerned with the mission of the district in which they are employed and that they strive for excellence.

Both Carroll and Lazovic refer to self appraisal and peer review as systems pertaining to the university faculty member; what is interesting is that both are formative in nature. If education is to improve and better people are to seek out the role of the public school teacher, then we must accord them the same professional privileges we have given university faculty, leaving peer review and self appraisal as the only systems of teacher evaluation that professional teachers can accept.

Neither peer review nor self appraisal has much of a future on their own as a public school teacher evaluation system -- not given the inherent acceptance of punitive or rewarding systems by the American public. The combination of both is intriguing since the development of peer-mediated self appraisal would provide the individual teacher with an evaluation system that they would generate and be responsible for, the purpose of which would be the improvement of their own teaching performance. That system could reflect board goals, district goals and individual building goals. It could take into account the items mentioned both by Carroll and Lazovic and it should help protect a teacher from biased evaluators. It should also provide for improved quality of instruction and have a clear objectivity check through peer review. This kind of evaluation could be done frequently enough to bring about updates of teaching materials, strategies and techniques.

The concept of peer-mediated self appraisal might allow differing political groups to agree on two things -- teachers can be evaluated and evaluation is not necessarily evil. If this concept were tried and developed fully, I believe I would see a great deal more collaboration on teacher evaluation from the several diverse conflicting groups now concerned with the improvement of teaching performance through evaluation. Any system that can succeed must be centered on the improvement of teaching performance -- in

essence, the evaluation of a process. Therefore, formative or process evaluation systems that have been developed and functioned efficiently in program evaluation should be used for theoretical background to examine how we might deal with teacher evaluation systems that are, in essence, formative. From that base, we should be able to develop a teacher evaluation system that is both objective and combines formative and summative components in ways that are congruent and acceptable to those being evaluated.

10. THE NATURE OF PEER-MEDIATED SELF APPRAISAL

Developing a teacher evaluation system is time consuming, costly, difficult to accomplish and further complicated by the fact that it must satisfy teachers, administrators, the community and the school board. It should include all employees and connect to the inservice training program. Peer-mediated self appraisal (PMSA) accomplishes all of this without the problems of earlier models. It is both formative and summative, satisfying the demands of teachers, administrators and evaluators. Figure 2 reveals seven connected elements that provide a smooth running, efficient and workable teacher evaluation system.

New and Beginning Teachers

PMSA's first component concerns new and beginning teachers. Summative in nature, it stresses product evaluation and enables administrators to decide if new and beginning teachers can teach to the standard competency level established by the district.

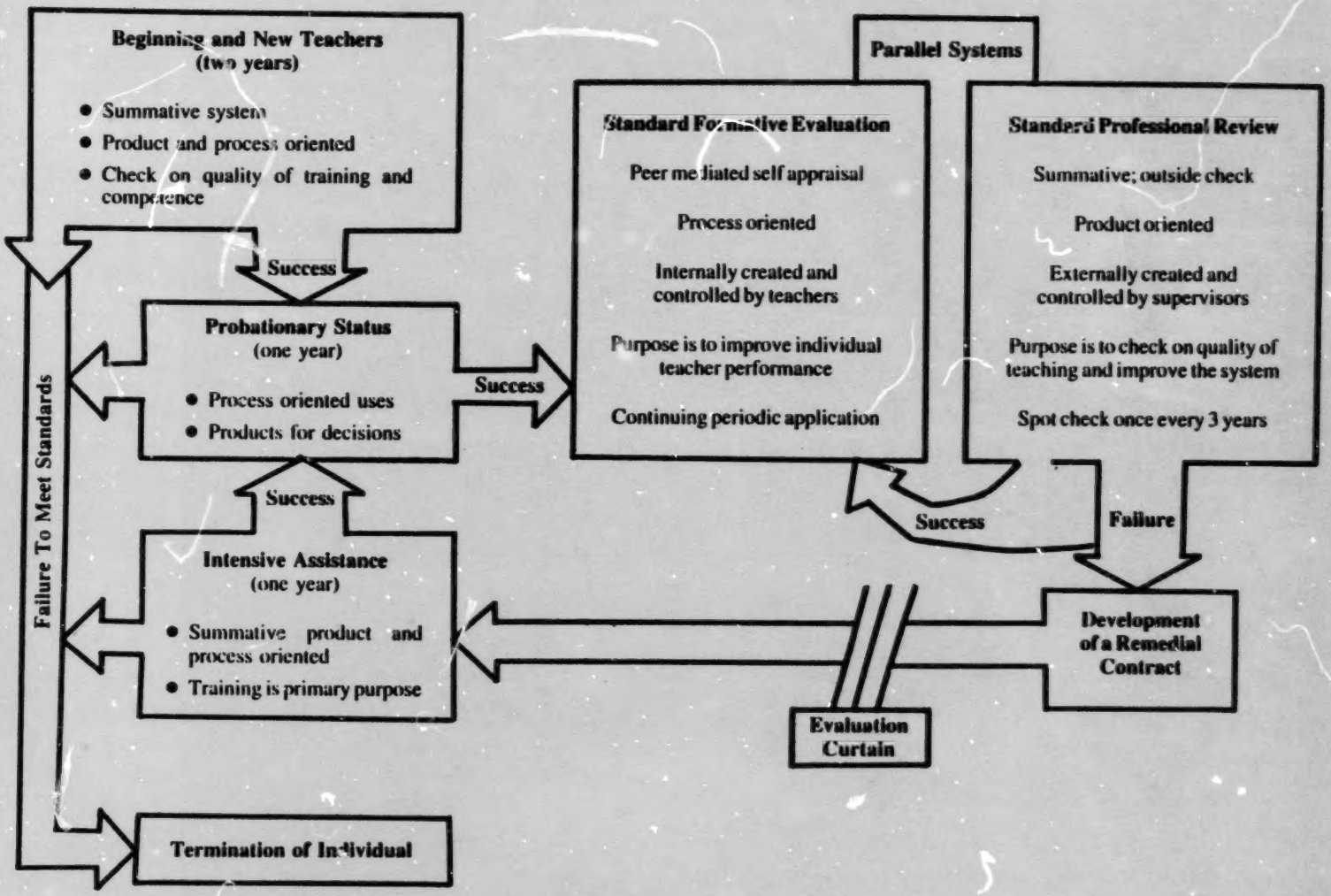
New and beginning teachers have an opportunity to

1. Show their ability to teach to the standard level
2. Show their ability to conduct an adequate self appraisal that they can then use to improve the quality of their teaching
3. Show that they can use peer review to improve their teaching
4. Determine if their training has been adequate to allow them to teach at the level of the district standards
5. Receive instruction in effective teaching techniques and strategies

PMSA should be considered an extension of preservice training. Periodically, evaluators carefully monitor performance and teachers review help in learning new skills.

If new or beginning teachers meet the criterion level established by the district over a period of two years, they enter the second component, probationary status. If they fail to meet the criterion level, they are terminated. The probationary status component is summative in nature, but less so than the beginning and new teacher component. Two kinds of teachers enter -- new teachers and teachers who were previously tenured but then placed on intensive assistance, which they have successfully completed. Regardless of how they get there, all people placed in probationary status are treated equally. They have the opportunity to show that they can successfully use the standard

Figure 2.


Beginning and New Teachers
(two years)
• Summative system
• Product and process oriented
• Check on quality of training and competence
Success
Probationary Status
(one year)
• Process oriented uses
• Products for decisions
Success
Intensive Assistance
(one year)
• Summative product and process oriented
• Training is primary purpose
Success
Failure To Meet Standards
Termination of Individual
Parallel Systems
Standard Formative Evaluation
Peer mediated self appraisal
Process oriented
Internally created and controlled by teachers
Purpose is to improve individual teacher performance
Continuing periodic application
Standard Professional Review
Summative; outside check
Product oriented
Externally created and controlled by supervisors
Purpose is to check on quality of teaching and improve the system
Spot check once every 3 years
Success
Failure
Development of a Remedial Contract
Evaluation Curtain

formative evaluation system and that they can teach to the standard set by the district.

Administrators must make major decisions about probationary status. After one year they decide whether teachers can be placed in the standard formative evaluation system and be tenured, or whether they must receive intensive assistance. Administrators must decide if people who have received intensive assistance, been placed on probationary status and then fail to achieve tenure must be sent back to intensive assistance or terminated. They must also decide if beginning teachers who fail the probationary period are placed in intensive assistance or recommended for termination. Only those teachers who successfully complete probation are recommended for tenure and placed on the standard formative evaluation system.

Standard Formative Evaluation System

The next component, standard formative evaluation system, deals with the process and strategy of teaching. Its purpose is solely to improve individual teacher performance. Teachers own the evaluation strategy and the results of self evaluation or peer review. Inservice training provides teachers with the information, skills and techniques to use peer review information to individualize and improve their teaching. Training is provided by a board made up solely of teachers and held in teacher centers run with funds and resources provided by the district.

Under this system, tenured teachers are professionally obligated to undergo self evaluation annually and to use peer review periodically. Teachers must keep their own records, proving that self evaluation and peer review have occurred. These records (but not the ones that show individual correction procedures) should be presented to a supervisor during the next component, professional review, which should occur at least once every three years.

Prior to the standard professional review, teachers and their employers should have developed a contract identifying the obligations of each. The school system must provide teachers with a job description and a list of the district's minimum standards of excellence in teaching, in addition to criteria for judging competence in the professional review, intensive assistance and probationary status components. All teachers must know the expectations for each component, and that their professional obligation includes self evaluation and peer review in the standard formative evaluation component.

Failure of a teacher to meet the terms of the contract during the three-year standard formative evaluation period may place that teacher on intensive assistance after the standard professional review. Failure of the district to meet the terms of the contract

automatically places the teacher on another three-year cycle of standard formative evaluation.

Standard Professional Review

The next component, standard professional review, will satisfy the board, administration, and state that all tenured teachers are teaching at least to the standard level expected by the district. Standard professional review, a summative system, uses product data agreed on by teachers, evaluators and administration. Data might include student achievement scores and records kept by the teacher during the standard formative and evaluation component, for example. In practice, the standard professional review compares the teacher data and performance with the district's standards and expectations and checks to see if the teacher has complied with standards. The teacher's supervisor conducts the standard professional review. It is imperative that the review utilize low inference measures such as structured observational schedules where data collected is a record of behavior, rather than any form of high inference measure (Soar et al., 1983). Both supervisor and teacher have the same information required to complete this spot check of teaching quality. A contract explains what each party must do and what data will be used to describe teaching quality criteria. The contract takes into account special circumstances like special education classes or unusual teaching assignments. The supervisor/evaluator owns this data but must share it with the teacher. Due process and equal consideration conditions are stated clearly and a teacher's property rights are protected. The review occurs at the end of the third year of a teacher's tenure or every three years thereafter and should occur during a specified period of time.

Inappropriate or inadequate teacher behavior may be dealt with more often than once every three years. For example, if the physical education teacher is unable to keep children from running the halls during gym class, the principal must feel free to inform the teacher that the rest of the building should not have to tolerate that kind of distraction and the teacher should have his charges under control in the classroom in the next five minutes, not three years from now during the professional review. This model does not intend to remake the whole system. Its purpose is to allow, at any one point in time, two-thirds of the faculty to control their own evaluation -- an evaluation that is formative in nature and designed to improve the individual's teaching performance without threat or intimidation.

At the end of the standard professional review, administration must decide whether to maintain teachers on the standard formative on system or to place them on intensive assistance. Administrators make decisions according to the terms of the evaluation contract. If teachers comply with all the terms of the contract and provide evidence to support compliance, then they

should remain on the formative system. Failure to comply with even one condition of the contract should be sufficient prima facie evidence to place teachers on intensive assistance. If the district or the district evaluator meets all the terms in the contract, they have the freedom to judge the teachers' next placement. Failure of the district or the evaluator to comply with even one term of the contract allows the teacher to continue on the formative system for three more years or until the next professional review.

Remedial Contract and Evaluation Curtain

Teachers who fail to pass the standard professional review are placed on intensive assistance. Two things occur before they move to intensive assistance: drawing an evaluation curtain and developing a remedial contract. The contract should specify what behaviors need correction, the terms of assistance and the evaluation criteria. The test of each part of the contract is twofold: (1) Does it clearly specify desired behaviors that are outlined in the teaching standards of the district? (2) Is each contract term reasonable? Is this a reasonable behavior to request of the teacher?

Following the development of a remedial contract, an evaluation curtain should be pulled over all previous existing data, including that data used to judge placement of the teacher on intensive assistance. The curtain announces that this teacher's evaluation starts again at zero. The teacher receives assistance to meet the standard teaching competence required by the district. Administration considers only the requirements stated in the remedial contract, along with any data gathered during intensive assistance, to determine whether to place the teacher on probation or to terminate.

Intensive Assistance

During intensive assistance the teacher receives inservice training or counseling in an effort to meet the standard teaching level of the district. This is a summative system and all data is owned by the evaluator. The teacher may take advantage of outside assistance, university training or whatever resources the district has available. The teacher may be required to teach, or learn, or both.

In this component the teachers receive written orders that they must obey during the required behavior change. All written orders must be reasonable and valid, and tied to the standards for the district. Each order must have a specified timeline. If the teacher fails to comply with written orders that are reasonable, valid, timely and just, then the teacher should be terminated on the grounds of insubordination or failure to follow a reasonable

order. If the teacher meets all the terms of the contract, follows the orders and meets district standards, he or she is placed on probation and given an opportunity to demonstrate competence under the standard formative evaluation system, while being observed and taught how to self evaluate and use peer review.

Assumptions

Several assumptions presented in the above model need clarifying. First, the district must identify and adopt as policy clear standards for teaching. Much of this work has been done during the past few years as a product of work in teacher evaluation by hundreds of districts. This work has provided job description statements of exemplary and standard teaching, behavior checklists, assessment measures, examples of board policy, and inservice training programs in sufficient numbers to put together a model system of standards in teaching excellence.

Second, the district will adopt and use objective measures in the professional review component of this program. The problem here is that the supervisors/evaluators (principals) don't like to use objective measures, calling them inadequate. Thus an implied third assumption is that principals will receive training in the use of objective measures and a major part of their evaluation will consist of examining how well they evaluate others.

The fourth assumption concerns the district's willingness to staff an inservice training program and fund a teacher evaluation program sufficiently to make it work.

The fifth and final assumption is the most difficult to deal with: that the administration, board, public, and even the teachers themselves are willing to accept that the position of "teacher" in this society is a professional position of importance. To accept this assumption implies that the society accepts that "teachers" are or should be (1) professionals who can judge their own worth, (2) professionals who strive for excellence in teaching, (3) inherently "good" people of worth to the community, (4) people who can and should assist in their own evaluation and the evaluation of their peers and (5) provided with opportunities for continuous upgrading of their skills that are supported by public funds.

REFERENCES

Abbott, Max G. "Evaluating School Administrators: The Scope and Nature of Administrative Performance," The Evaluation of Administrative Performance: Parameters, Problems and Practices, ed. by William J. Gephart, Robert B. Ingle and W. James Potter. Bloomington, Ind.: Phi Delta Kappa, 1975.

Anderson, Scarvia B., Samuel Ball and Richard T. Murphy. Encyclopedia of Educational Evaluation. San Francisco: Jossey-Bass, 1975.

Bhaerman, Robert D. "Merit Pay? No!" National Elementary Principal, vol. 52, 1973.

Brinks, James T. "Is There Merit in Merit Increases?" Personnel Administrator, vol. 25, 1980.

Carroll, J. Gregory. "Faculty Self-Evaluation," Handbook of Teacher Evaluation, ed. by Jason Millman. Beverly Hills, Calif.: Sage Publications, 1981.

Doremus, Richard R. "What Ever Happened to Kalamazoo's Merit Pay Plan?" Phi Delta Kappan, vol. 63, 1982.

Glass, Gene V. "A Paradox About Excellence of Schools and the People in Them," Educational Researcher, vol. 4, 1975.

Johnson, Susan Moore. "Performance Based Staff Layoffs in the Public Schools: Implementation and Outcomes," Harvard Educational Review, vol. 50, 1980.

Knapp, Michael S. "Teacher Evaluation Practices Within Schools: An Overview of the Literature and Issues From the Practitioner's Point of View," CEDR Quarterly, vol. 15, 1982.

Lazovic, Grace French. "Peer Review: Documentary Evidence in the Evaluation of Teaching," Handbook of Teacher Evaluation, ed. by Jason Millman. Beverly Hills, Calif.: Sage Publications, 1981.

Monk, James A. "My Sad Conclusion: Board Turnover Precludes Merit Pay for Administrators," The Executive Educator, vol. 5, 1983.

Popham, W. James. Educational Evaluation. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Redfern, G.B. Evaluating Teachers and Administrators: A Performance Objectives Approach. Boulder, Colo.: Westview Press, 1980.

Rice, Berkeley. "Space Lab Encounters," Psychology Today, vol. 17, 1983.

Robinson, Glen E. Merit Pay for Teachers. Arlington, Va.: Educational Research Service, 1979.

Robinson, Glen E. "Paying Teachers for Performance and Productivity: Learning From Experience," speech delivered April 11, 1983 to the 100th Annual Conference of Virginia Division Superintendents, Roanoke, Va.

Scriven, Michael. "A Different Approach to Teacher Evaluation," Planning for the Evaluation of Teaching, ed. by Willard R. Duckett. Bloomington, Ind.: Phi Delta Kappa, 1979.

Scriven, Michael. "The Methodology of Evaluation," Educational Evaluation: Theory and Practice, ed. by Blaine R. Worthen and James R. Sanders. Belmont, Calif.: Wadsworth, 1973.

Scriven, Michael. "Summative Teacher Evaluation," Handbook of Teacher Evaluation, ed. by Jason Millman. Beverly Hills, Calif.: Sage Publications, 1981.

Silverman, Buddy Robert. "Why the Merit Pay System Failed in the Federal Government," Personnel Journal, vol. 62, 1983.

Soar, Robert S., Donald M. Medley and Homer Coker. "Teacher Evaluation: A Critique of Currently Used Methods," Measurement Based Evaluation of Teacher Performance. Longman (in press).